Discriminative Regularization for Latent Variable Models with Applications to Electrocardiography

Poster #53

Andrew C. Miller

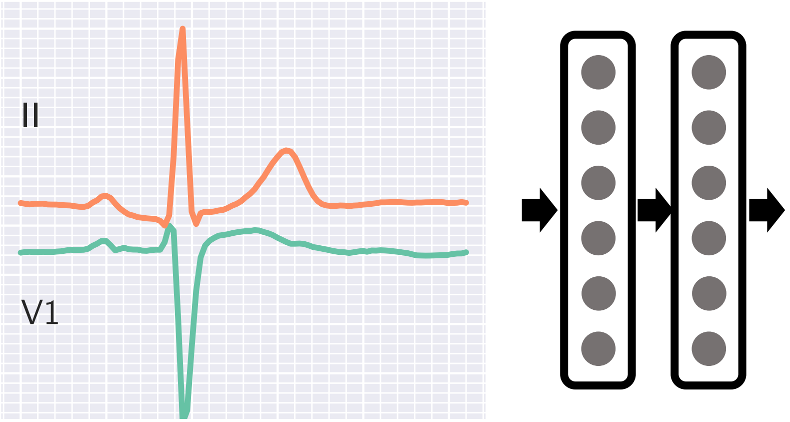
with Ziad Obermeyer, John P. Cunningham, and Sendhil Mullainathan

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Berkeley School of Public Health

CHICAGO BOOTH The University of Chicago Booth School of Business

Black-Box Predictors



$$Pr(\mathbf{y} = 1 \,|\, \mathbf{x})$$

(e.g. revascularization, future afib diagnosis, high troponin)

$$m(\mathbf{x})$$

discriminative/ predictive model

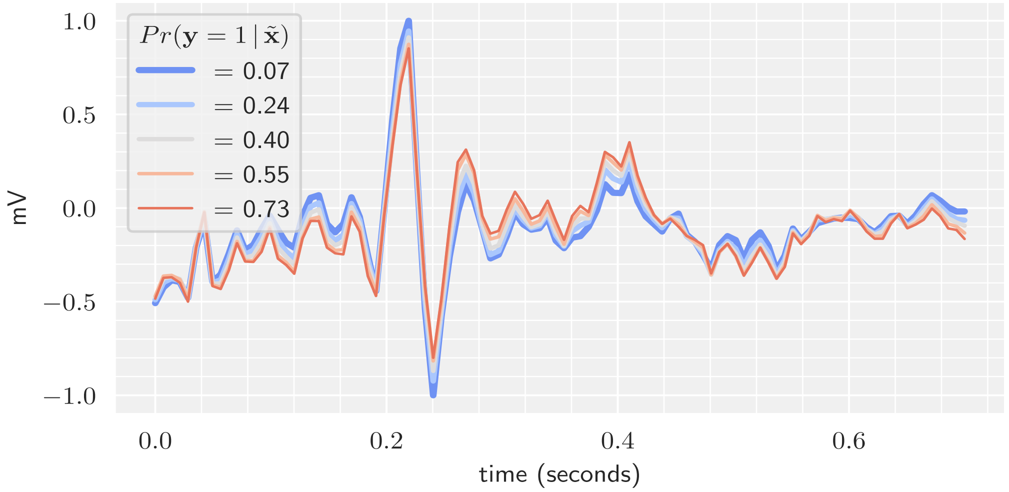
Accurately assess (some) disease risk.

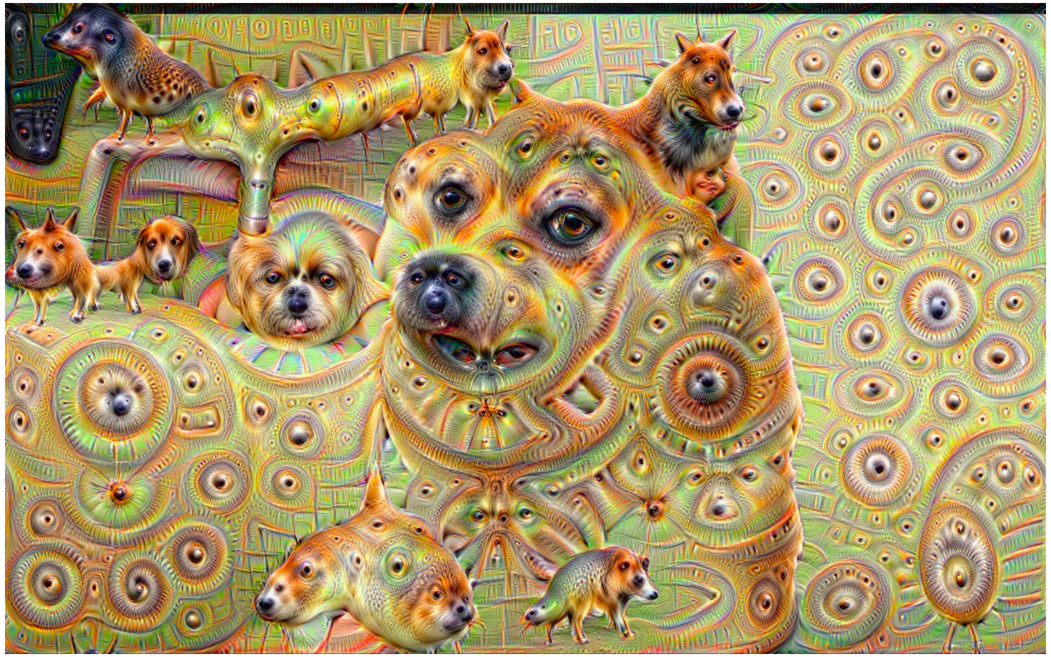
What is the algorithm "seeing"?

Gradient-based Explanations

Follow naive gradient (e.g. saliency maps)

$$\nabla_{\mathbf{x}} m(\mathbf{x})$$





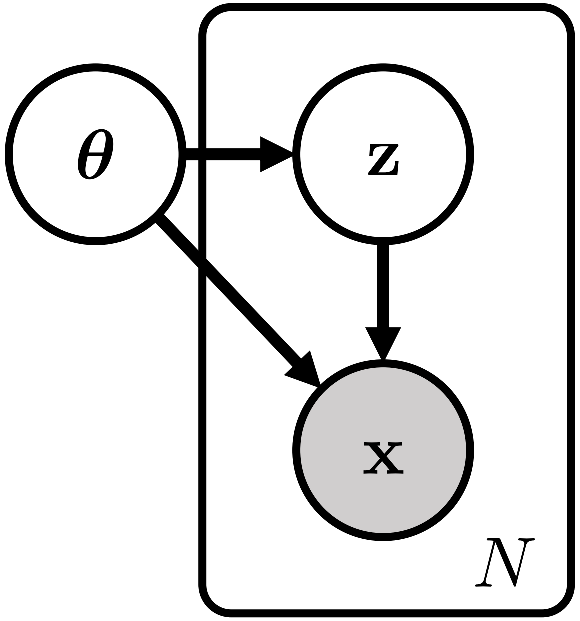
Gradient is a local object; ignores global structure in $\mathbf{x} \in \mathcal{X}$.

Want to explore *realistic predictive features.* Need to *model* the structure in x

https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Generative Latent Variable Models

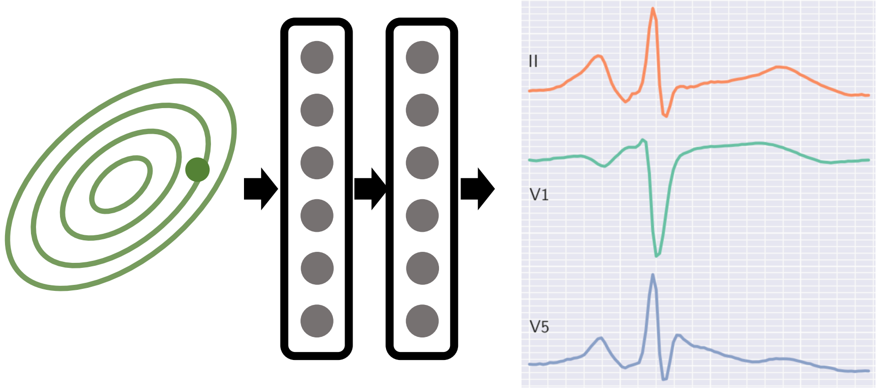
Global variables



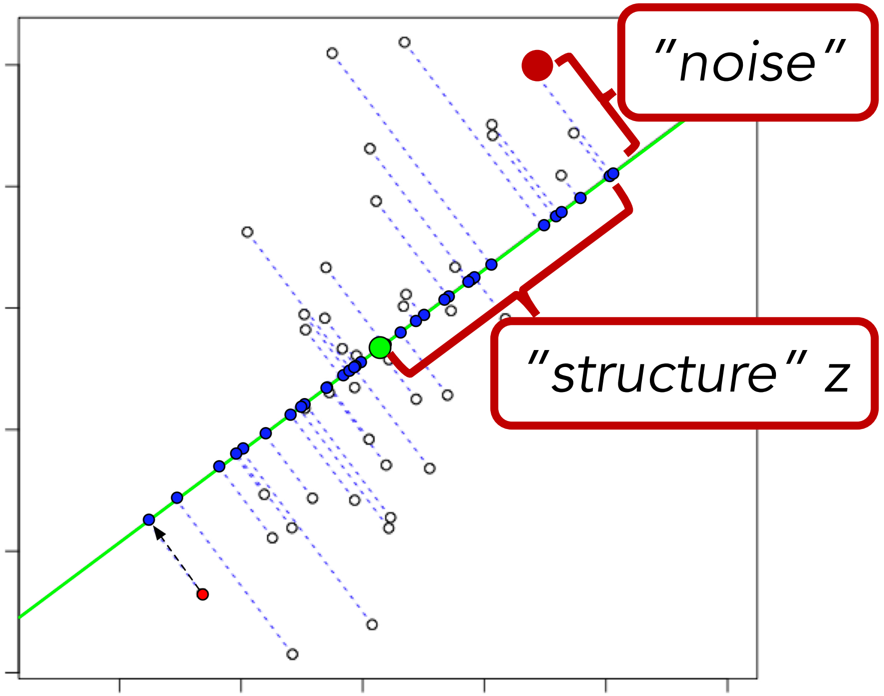
Per-observation (low-dimensional) latent variable

Observed data (e.g. Image, EKG)

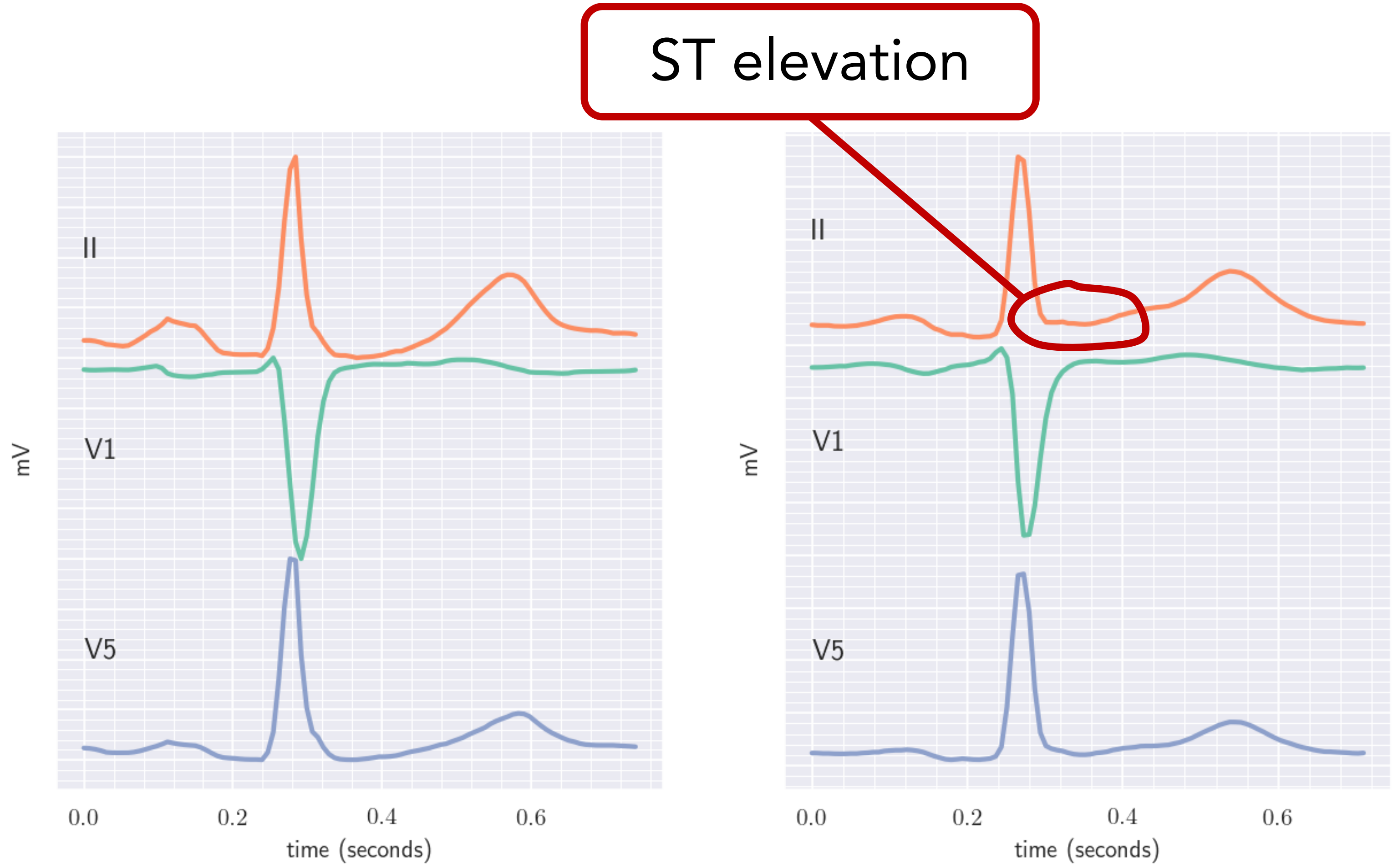
Deep Generative Models:



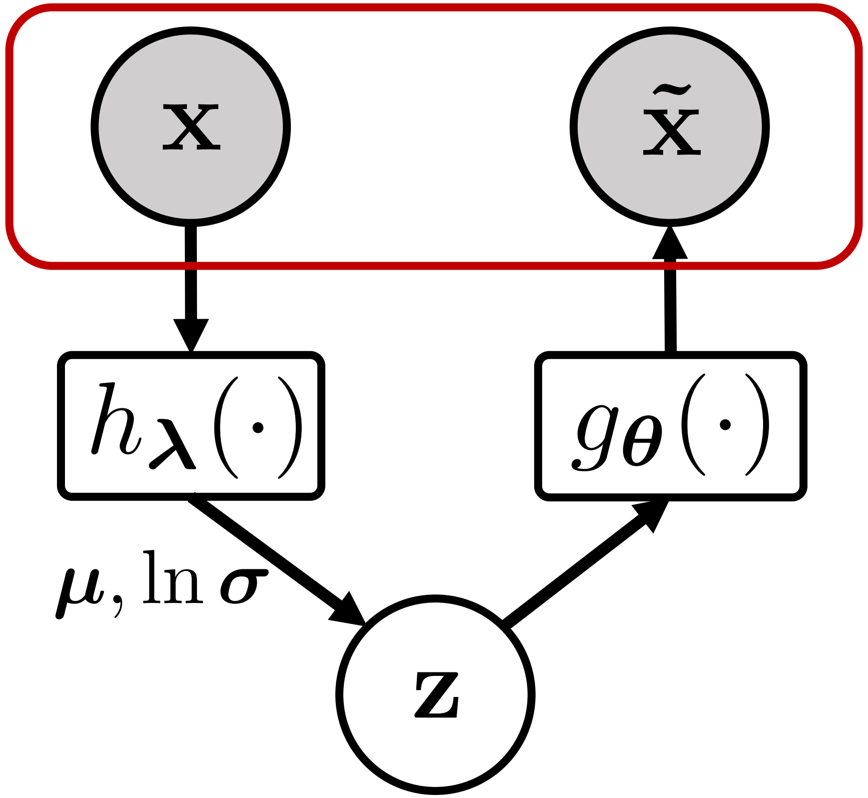
Structure vs. Noise
e.g. probabilistic PCA



Unsupervised Models Drawbacks



maximum likelihood preserves features with high variability – like non-linear PCA



How to make **z** sensitive to subtle features?

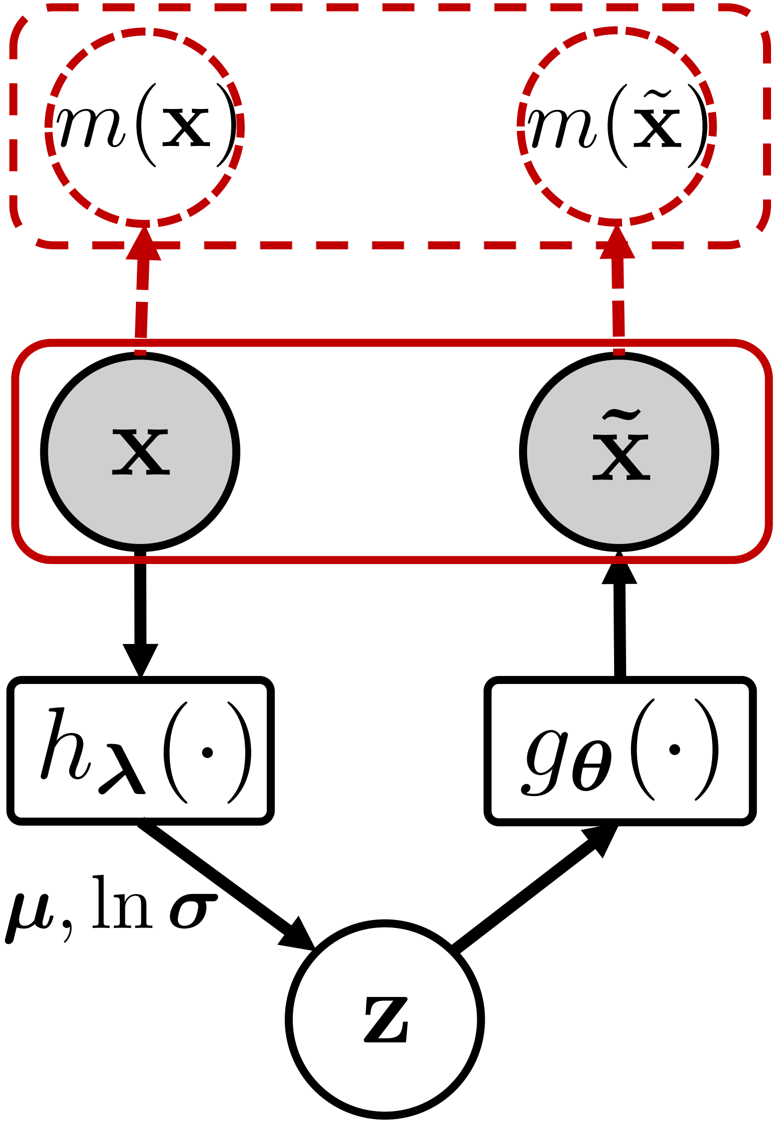
Discriminatively Regularized VAE (DR-VAE)

DR-VAE Objective

$$\mathcal{L}_{discrim}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = D_{KL}(m(\mathbf{x}) \,||\, m(\tilde{\mathbf{x}}))$$

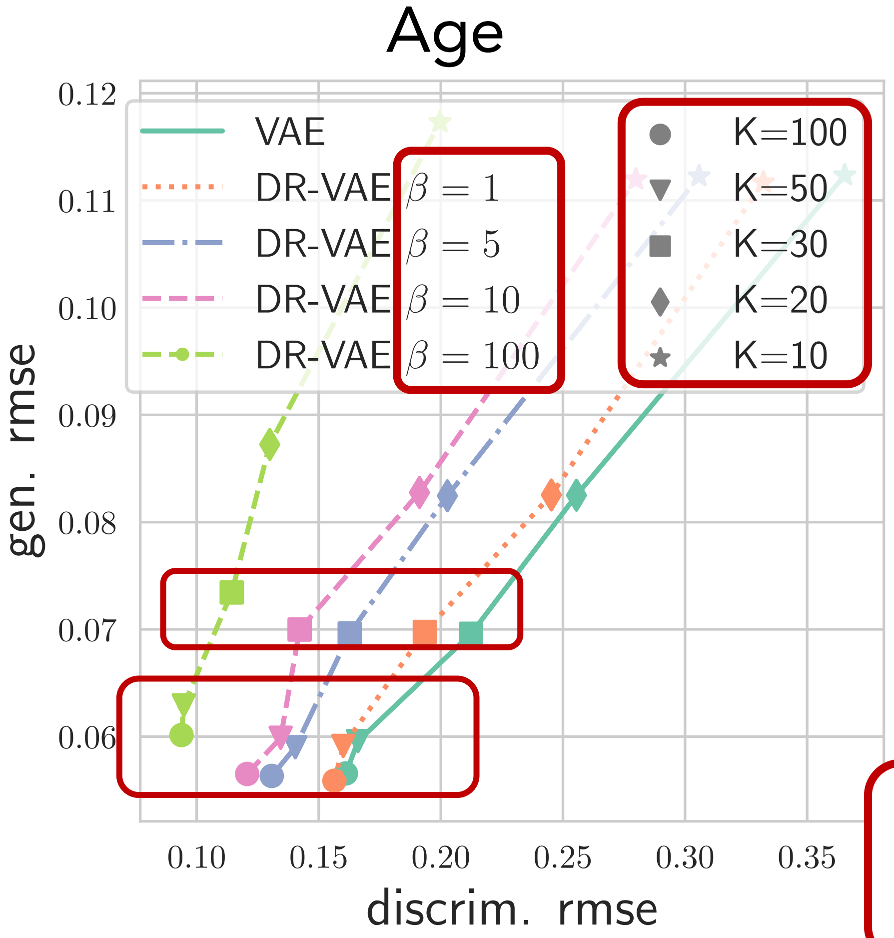
$$\mathcal{L}_{dr\text{-}vae}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathcal{L}_{elbo}(\boldsymbol{\theta}, \boldsymbol{\lambda}) + \beta \cdot \mathcal{L}_{discrim}(\boldsymbol{\theta}, \boldsymbol{\lambda})$$

typical VAE obj.   discrim. regularizer

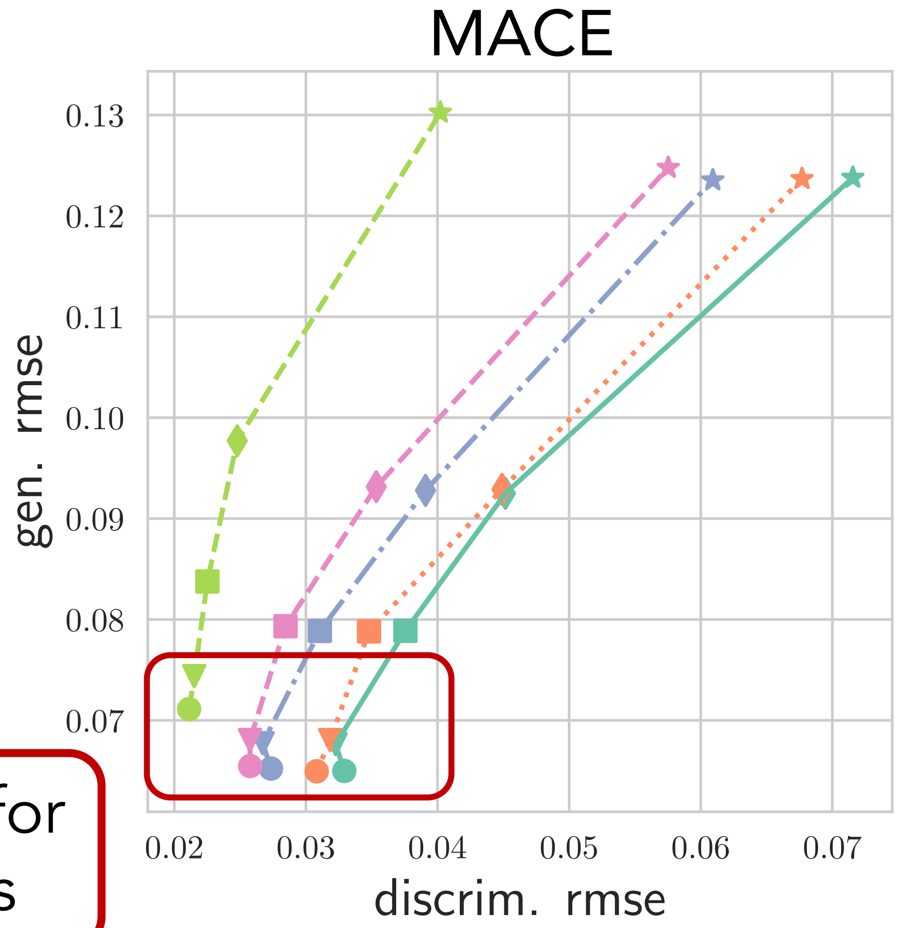


EKGs: Discriminative-Generative Tradeoff

$\mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})^2]$ (gen error) vs $\mathbb{E}[(m(\mathbf{x}) - m(\bar{\mathbf{x}}))^2]$ (discriminative error)



Little trade-off for flexible DGMs

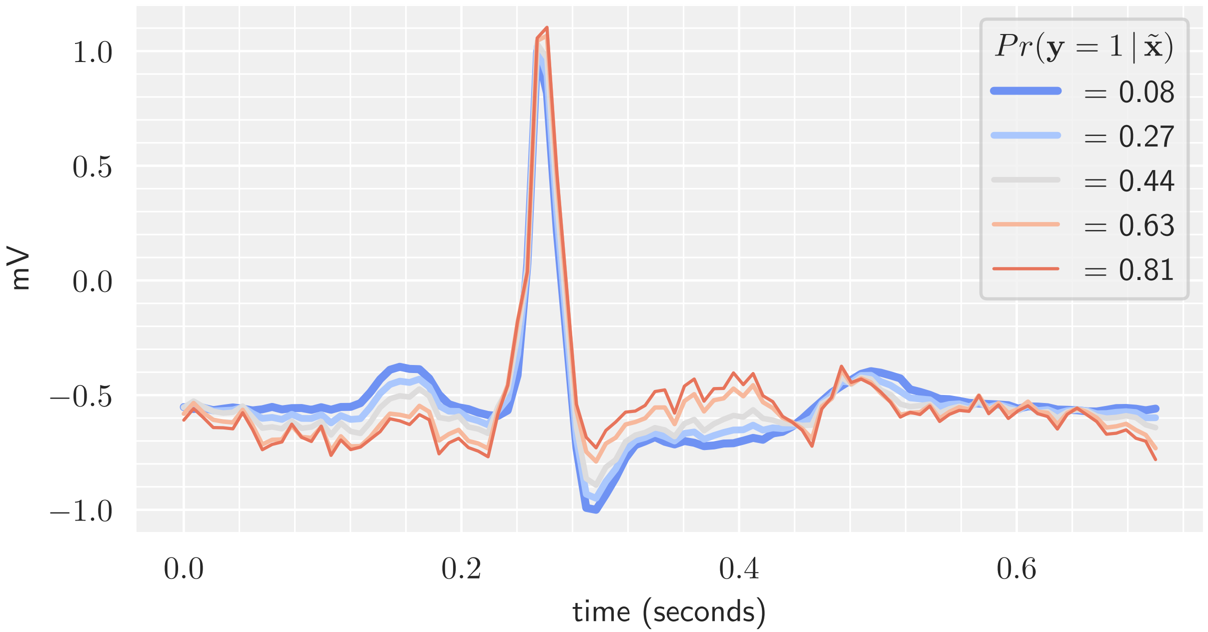


Model-Morphs

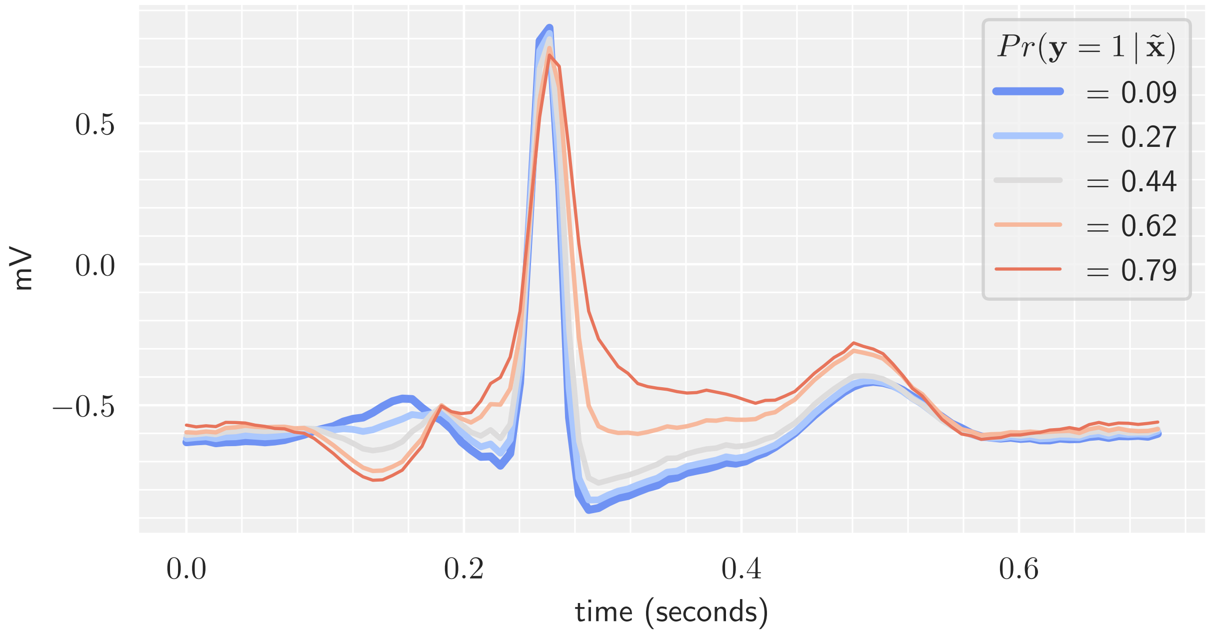
Generate "morphing trajectory", follow model gradient in latent space

$$\mathbf{z}^{(t+1)} \leftarrow \mathbf{z}^{(t)} + \delta \cdot \frac{\partial m}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}(\mathbf{z}^{(t)})$$

$$\tilde{\mathbf{x}}^{(t+1)} \leftarrow g_{\boldsymbol{\theta}}(\mathbf{z}^{(t+1)})$$



Model free



DR-VAE

More details at the poster!

- Technical details, application, and quantitative comparisons

Poster #53
(Wednesday night)